

Universidade de Pernambuco

Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

Proposta de Dissertação de Mestrado

Área: Inteligência Computacional

Título: Aprendizado Federado para Saúde Digital

Orientador – Bruno José Torres Fernandes (bjtf@ecomp.poli.br)

Descrição – O avanço da tecnologia e a coleta massiva de dados têm impulsionado o desenvolvimento de algoritmos de aprendizado de máquina cada vez mais complexos e eficientes [1]. No entanto, muitas vezes esses dados são distribuídos em diferentes locais, o que torna difícil a tarefa de treinar um modelo com essas informações [2]. Além disso, o compartilhamento desses dados pode violar a privacidade das pessoas ou empresas envolvidas [3].

O aprendizado federado é um paradigma recente [4] que permite o treinamento de modelos de aprendizado de máquina em conjuntos de dados distribuídos e descentralizados, sem a necessidade de compartilhar os dados diretamente entre os participantes. O aprendizado federado permite então que múltiplas entidades (por exemplo, dispositivos móveis, sensores, servidores em diferentes locais) colaborem para treinar um modelo sem compartilhar seus dados brutos. Esta abordagem promove a colaboração entre diferentes entidades, ao mesmo tempo em que protege a privacidade dos dados e minimiza os custos associados à comunicação e ao armazenamento centralizado e pode ser utilizada em diversos cenários, desde segurança urbana até cidades inteligentes [5].

Por outro lado, a crescente digitalização no setor de saúde gerou uma vasta quantidade de dados médicos provenientes de dispositivos wearables, exames de imagem e registros eletrônicos de saúde, os quais possuem grande potencial para desenvolvimento de modelos baseados em aprendizado de máquina. No entanto, a centralização desses dados traz desafios significativos relacionados à privacidade, segurança e conformidade regulatória, como as diretrizes da LGPD e GDPR. Dessa forma, o aprendizado federado surge como uma solução promissora, permitindo o treinamento de modelos de aprendizado de máquina em dados distribuídos, sem a necessidade de compartilhamento de dados brutos entre instituições ou dispositivos. Isso preserva a privacidade dos pacientes enquanto promove colaborações em larga escala no setor de saúde [6].

Esta proposta de pesquisa visa desenvolver e avaliar uma arquitetura de aprendizado federado para análise de dados médicos distribuídos, com foco em imagens médicas e sensores wearables. O objetivo principal é treinar modelos robustos e precisos que respeitem a privacidade dos pacientes e atendam aos requisitos de aplicações práticas na saúde digital.

O trabalho contribuirá para avanços na área de aprendizado de máquina aplicado à saúde, oferecendo uma solução viável e segura para análise de dados médicos distribuídos. A proposta também se alinha aos Objetivos de Desenvolvimento Sustentável (ODS), incluindo saúde e bem-estar (ODS 3) e inovação e infraestrutura responsável (ODS 9).

Referências Bibliográficas

1. NGUYEN, G. T. et al. Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review*, v. 52, p. 77–124, 2019.
2. TULADHAR, A. et al. Building machine learning models without sharing patient data: A simulation-based analysis of distributed learning by ensembling. *Journal of biomedical informatics*, p. 103424, 2020.
3. ZHANG, T.; HE, Z.; LEE, R. B. Privacy-preserving machine learning through data obfuscation. *ArXiv*, abs/1807.01860, 2018.
4. MCMAHAN, H. B. et al. Communication-efficient learning of deep networks from decentralized data. In: *International Conference on Artificial Intelligence and Statistics*. [S.l.: s.n.], 2016.
5. ALEDHARI, M. et al. Federated learning: A survey on enabling technologies, protocols, and applications. *IEEE Access*, v. 8, p. 140699–140725, 2020.
6. Mohammed Aledhari et al., "Federated Learning: A Survey on Enabling Technologies, Protocols, and Applications," *IEEE Access*, 2020.